DOCUMENT RESUME

ED 160 622 TM 007 546

AUTHOR Lawrence, Jeanette A.
TITLE Moral Judgment Intervention Studies Using the Defining Issues Test.
PUB DATE [77]
NOTE 34p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Developmental Stages; *Intervention; *Moral Development; Research Design; *Research Reviews (Publications); *Test Validity
IDENTIFIERS Construct Validity; *Defining Issues Test

ABSTRACT

Fourteen moral judgment intervention studies, representing all those presently known to the author, were reviewed and evaluated. All used Rest's Defining Issues Test (DIT) to evaluate Kohlberg's stage theory of moral development. The DIT had been administered in programs for social studies and civics, general psychological development, and moral education. This review addressed two questions: Is it possible to stimulate the development of moral judgment in an educational environment?; and Does the DIT assess such moral judgment changes? Results showed that significant changes on the DIT were related to duration and type of intervention and degree of experimental control over treatment variables. The studies provided independent support for the ability of the instrument to index upward movement after intervention programs designed to influence development. Three suggestions emerged from this review: (1) interventions should be tailored to fit moral judgment theoretical concepts; (2) the measurement instrument should be a good fit for the treatment; and (3) future research should focus on moral and judgmental aspects of moral judgment programs. Table I presents an overview of the fourteen studies including their sample sizes, characteristics, design focus, instrumentation and testing, and evidence of change on the DIT. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED160622

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jeanette A. Lawrence

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

MORAL JUDGMENT INTERVENTION STUDIES

USING THE DEFINING ISSUES TEST.

Jeanette A. Lawrence,

University of Minnesota

Running head: Moral judgment interventions

Requests for reprints should be sent to the Department of Social, Philosophical and Psychological Foundations of Education, 330 Burton Hall, University of Minnesota, Minneapolis, Minnesota, 55455. I wish to acknowledge financial support from the National Institute of Mental Health, Grant #R01 MH27861.

TM007 546

Abstract

Fourteen moral judgment intervention studies were reviewed and evaluated. All used Rest's (1974) Defining Issues Test as an index of moral judgment development. Results showed that significant changes on the Defining Issues Test were related to duration and type of intervention and degree of experimental control over treatment variables. The studies provided independent support for the ability of the instrument to index upward movement after intervention programs designed to influence development. Half the studies produced significant developmental change.

MORAL JUDGMENT INTERVENTION STUDIES
USING THE DEFINING ISSUES TEST

Moral development intervention studies usually follow one of a few well-used models: moral dilemma discussions, modeling of higher levels of moral reasoning, or stimulation of general psychological (including moral) growth. They have traditionally used the Kohlberg interview and scoring techniques to index changes in their subjects' levels of moral thought. The field is in need of new directions for innovative, theory-related, rigorous research. What has been done had not been systematically reviewed prior to Lockwood (1977) and this present review.

Lockwood provided a much needed critique of interventions in the Values Clarification and Kohlberg traditions. This review discusses the cognitive-developmental approach to moral education programs by examining objectives, results and effectiveness of fourteen recent intervention studies most of which have not been published, and all of which use Rest's (1974) Defining Issues Test (DIT) as the index of moral judgment development. Some of them follow one of the usual patterns; others suggest possible ways to advance program-development, such as the tailoring of treatment to theoretical concepts of moral judgment and the measurement instrument (Panowitsch, 1975; Coder, 1975).

The studies were not selected for excellence, nor for commitment to a particular research interest, they represent all the intervention studies using the DIT presently known to its author. They are the work of a variety of researchers with divergent objectives and interests; they have varied sample characteristics and experimental designs. The group includes some that did not produce upward movement on the DIT, as well as some that obtained significant change. The critique of these studies as a body provides a basis

for drawing inferences about the characteristics of effective moral education programs, and the construct validity of the DIT.

THE DEFINING ISSUES TEST: AN INDEX OF MORAL JUDGMENT DEVELOPMENT

The DIT is an objective test of Kolhberg's stage theory of moral development, with Rest's (1976a) modifications of the stage characteristics. Although longitudinal and cross-sectional data have demonstrated the validity of the test as an index of the development of moral reasoning, to date no systematic review of intervention studies has been available. This review addresses that deficiency.

Essentially the test presents the subject with six Kohlberg dilemmas, and a set of seventy-two items, samples of stage-related reasoning to rate and rank as ways of solving the dilemmas. Distraction and check items are included to test for inconsistent or faking responses. The scoring system used in the studies is the Principled Thinking (P) Scale which indexes the number of stage 5A, 5B and 6 items ranked as most important in solving the moral problem. A new index, the D scale, is at present being tested; it indexes the decrease in the subject's use of lower stage items as well as the increase in P items, (Davison, Robbins, Swanson, 1977). The D scale also uses the subject's rating responses. Although both scales are sensitive to developmental change, the D scale seems to be more sensitive to upward change. By using the DIT to measure changes in moral judgment levels, the interventionist is implicitly, if not explicitly, adopting a developmental, and moral judgment perspective for his study. Results obtained are relevant in answering two important questions: "Is it possible to stimulate moral judgment development in an educational environment?" and "Does the DIT index such moral judgment changes?"

This review addresses those questions; examines the findings of the studies, discusses their relative contributions to the understanding of moral judgment development, and makes some general observations about interventions designed to facilitate the development of moral reasoning. In general, difficulties in the interpretation of the results seem to arise from: manipulation of too extensive multivariate treatments, inadequate experimental controls, and insufficient attention to theoretical considerations.

OVERVIEW OF THE STUDIES

Table I presents a summary overview of the studies, their sample sizes and characteristics, design and focus, instrumentation and testing, and evidence of change on the DIT. The studies have been grouped for convenience according to two criteria, duration, and type of treatment experience provided in the educational program. These programs were in the areas of social studies, psychological development, or moral development education.

TABLE I ABOUT HERE

SUMMARY OF OBJECTIVES, FINDINGS AND CHARACTERISTICS OF THE STUDIES.

A modeling study with a brief, one-lesson treatment.

Walker (1974) presented seventy eighth graders with written narratives modeling reasoning one stage above, one stage below, or at the subject's own predominant stage. This study replicates Turiel's (1966) method of "plus one modeling", but with the use of narrative modeling. Subjects at stage 2 and stage 4 were each grouped into three conditions and were exposed to written reasoning at one stage above (+1), one stage below (-1), or at their own stage. Both pro and con reasoning was given for a particluar solution.

Walker hypothesized that only "plus one" modeling would influence upward movement on the DIT, but found no confirmation for this. The plus-one group

did not change significantly. The assumption of this study was that one brief exposure to writte- reasoning would influence structured thought to produce upward change. Although it follows Turiel's (1966) well-known technique, the study does not produce convincing evidence that the organization of moral thought can be changed quickly, or with little environmental stimulation. The earlier studies in the genre likewise did not produce convincing results (Turiel, 1966; Tracy and Cross, 1973). This type of study is further complicated by the author's assumption that a stage-typing by predominant stage is accurate and stable (Rest, 1976(b)).

Social studies and civics programs.

The Minneapolis Police Report (Rest, Ahlgren, and Mackey, 1972) used instructional units to inform students about police activities, and to promote favorable attitudes to the police. Students in three junior high schools were exposed to instructional units in civics, social studies, and police activities, and practical experience with members of the police-force on the job. The comprehension of Police/Perception of Law Officers Scale (COPO/POLO) was used to measure attitude change to police. The DIT, COPO/POLO and three other traditional atttitude scales, were administered to experimental groups in pre-and posttests.

COPO/POLO showed a change in two of the three dimensions of perception of police: they were seen as more helpful, and more angry, i.e., involved rather than cold-blooded. There were significant increases on democratic, and political efficacy attitude scales. There were no significant pre-to

posttest changes

on the DIT's P scale; however, a decrease in stage 2 thinking [$t = -2.17$ ($p \leq .03$)], suggests that a re-analysis of the data using the D scale might reveal more positive trends.

The study is representative of a class of interventions in which a primary aim is addressed, and the moral judgment measure added to give developmental context and meaning to the findings. The authors' rationale for using the DIT was to provide a "conceptual framework by which subjects interpret and make decisions about social events" (p. 10). The treatment, however, seems to have been only indirectly related to such theoretical presuppositions, and was not specifically designed to promote change in conceptual levels of moral thought. The authors assumed a transfer effect from attitude change to changes in moral judgment levels.

Another study in the general class of social studies and civics education is Morrison, Toews, and Rest (1975) "Youth, Law and Morality" programs which tried to raise subjects' level of moral reasoning by systematic learning activities based on civil rights issues. The experimental groups were given instruction in understanding legal procedure and problems of civil rights. Students were given the opportunity to engage in oral discussions, case studies, and argument flow analysis. The control groups had a course in British history.

From pre- to posttest, and posttest to the 6 week's follow-up, there were no significant changes that could be attributed to the experimental treatment on any of the thirty-one dependent variables, including nine based on the DIT, i.e., stage, P, Meaningless & Anti-establishment responses. There were no significant differences between experimental and control groups on posttest or follow-up. The multiplicity of treatment factors and dependent variables suggests that the experimenters may have missed out by trying to cover too much in a new program with administration problems, and by over-testing their subjects.

General psychological development programs.

Six of the present studies are representative of a large number of interventions conducted according to the Deliberative Psychological (DPE) model, or similar models of programs for the promotion of general psychological development. Typically subjects are exposed to a diverse set of experiences either inside or outside the clsssroom. These experiences may include such events as discussion of developmental theories, log-keeping, empathy training and community-based practicums. The objective of this multitreatment type of study is to promote general psychological development. The relation of that objective to moral judgment is expressed in a specific aim to promote growth of moral reasoning, or by assumed transfer to that type of conceptual organization from other independent variables.

Erickson, Colby, Libbey and Lohman (1976) applied the DPE model in a within-school module in an English elective class and recommended the course for inclusion in English, Social, Family or Health studies curricula. The treatment included discussions, projects, questionnaires and teaching units designed to promote psychological growth in the areas of self-identity, relationships and conflicts, social perspective-taking and understanding of laws and social systems. Teachers were cast in the role of psychological educators.

The Loevinger/Wessler Sentence Completion Test and the DIT were used as developmental indices, an empathy scale was also used. Significant changes were obtained on the DIT[t (19) = 2.27,] $p \leq .02$, both for a one-tailed t test. There was no significant change on the empathy scale.

Since teaching units on Kohlberg's stages were included in the treatment, and since no control group was used for "natural development" comparison, there is no certainty that the obtained differences in the DIT scores represent genuine structural, developmental changes in moral reasoning. Although the authors

explain that the study was the first application of the DPE model to a junior high school sample, the model itself had been sufficiently piloted to warrant the inclusion of a control group and auxiliary measures other than the empathy scale.

In another study in the genre, empathy training was considered to be a potential influence on the general psychological development of college students (Hurt, 1975). Hurt introduced auxiliary testing with a Responding Skills instrument designed to measure the effectiveness of empathy and involvement training, and any co-variation of that variable with ego and moral developmental change shown on the Loevinger and DIT tests.

As he was unable to assign subjects randomly to treatments, Hurt assigned, at random, three intact sections of the course to experimental, active control, and inactive control groups. The experimental group was given counselling (empathy) training in a practicum, plus a discussion seminar, and the traditional educational psychology component, the active control received all but the intensive empathy training, and the inactive control received only the educational psychology component plus one brief session in empathy training.

On the Empathy Skill Test, Hurt obtained pre- to posttest gains for both the experimental and active control groups, $\underline{t}(14) = 7.7$, $\underline{p} < .001$ and $\underline{t}(18) = 2.6$, $p < .01$ respectively. No significant results over the experimental period were obtained on the Loevinger instrument, but on the DIT gains were significant for the experimental group (on a one-tailed t test), $\underline{t}(14) = 1.94$, $\underline{p} \leq .037$ and for the active control, $\underline{t}(18) = 1.8$, $\underline{p} \leq .045$. The inactive control group showed no changes.

Sprinthall and Bernier (1977) also used an auxiliary skill acquisition instrument (as well as self-reports) to attempt to monitor developmental influences. Their study extended the use of the DPE model to practicing teachers in an effort

to promote both personal and professional development. Their curriculum had two components: a) an intensive summer workshop which included large group teaching, and small group training in counselling, teaching and individualized instruction skills; and b) a twice-monthly practicum during the following quarter's teaching.

Results showed a pre- to posttest change on the Reflection of Feelings Scale, $t(17) = 5.71$, $p < .0001$. There were significant pre- to posttest gains on the DIT $t(17) = 2.19$, $p < .01$ (on a two-tailed t test) but not on the Loevinger measure. No control group was used, but seven teachers from another location were introduced as a "comparison" group. They were reported to have shown no developmental gains on the DIT over the experimental period, but no data are given.

Balfour (1975) used the "action reflection" aspect of the DPE model to develop a "Humanities Outreach Program" in three Minnesotan high schools: at Lindbergh, with both experimental and control groups; at Stillwater, with a control group only for the pretest; and at West, with no control group. The objective was to create opportunities for social exposure and real-life laboratory experiences for the development of students' senses of responsibility and personal competency. Evaluation on the DIT was supposed to give the achievement of these objectives developmental meaning by indexing generalized changes. There was the implicit assumption that the multiple experiences would accelerate conceptual moral growth in the subjects. Treatments for experimental groups included practicums in community involvement, seminar discussions and field trips.

No significant pre- to pre-test change on the DIT was found in either of the two Lindbergh or the Stillwater groups. The West group showed significant changes, $t(11) = 1.84$, $p \leq .046$ and the combined experimental groups showed significant gains $t(53) = 2.01$, $p \leq .050$. The author justifies the combination

of the experimental groups on the ground that they had experienced the same practicum component, however, they were different on the pretest: the Lindbergh experimental group was significantly higher than the experimental groups at West (at the .030 level), and Stillwater (at the .034 level). The DIT dependent measure seems to have been chosen for economy. Moral judgment may not have been the most suitable measure for this study. To use it assumes transfer from practical experience to moral judgment development.

In another general psychological development intervention by Kickbusch and his associates (Allen, 1976), "Confluent Education" was described as the integration of affective and cognitive components of the school curriculum, and in that sense, it resembles the DPE model. With multifaceted treatments, and special emphasis on the influence of teaching styles, it was a deliberative attempt to promote psychological growth and integration. Experimental groups were exposed to confluent education, and controls to traditional teaching methods. A second control group from another course, World History, provided another control. Moral reasoning components were: a semester each in values clarification classes, and Kohlberg-type discussions of dilemmas. The authors also assumed values issues would emerge from the social studies components. No significant pre- to posttest gains were obtained on the DIT.

The authors claimed that the DIT was not a suitable measure for their study because the level of understanding necessary was too high for their ninth graders (Allen, 1976). Allen claims that the application of the consistency check (Rest, 1974) showed that inconsistency was the norm for their sample, and that the subjects did not understand the tasks. He attributed the absence of any significant relation between the DIT scoring and the multiple independent variables, at least in part, to test difficulty. However other ninth grade samples have been able to complete the test adequately (Rest, 1974),

and the absence of pre- to posttest changes may have been due to the multiplicity and confounding of treatments and over-testing, twenty dependent measures having been used. The multiplicity of dependent variables, in this large-scale intervention may have obscured specific effects. Such programs may be more effective with smaller experimental design and intensive piloting.

In the Sierra Project at University of California, Irvine, the intervention involved the total college freshman experience (Whiteley, 1976; Nelson, 1976). The program was designed to increase personal growth by the integration of college courses with informal educational experiences in the residential hall's community life-style. There were three groups: experimental, placebo control, and randomly selected control group. Experimental Sierra Hall, and placebo Lago Hall subjects were randomly assigned from a pool of 138 students requesting admission to the Sierra experimental program. The DIT and six other measures were used for pre- and posttesting, including the Kohlberg interview and Loevinger Sentence-Completion Test. During the eight months, several courses included instruction in moral and developmental theories, in particular a "Moral Development and Just Communities" course. Added to the pre- and posttests were intensive case study assessments on the experimental group subjects, and topical assessment of the curriculum modules. Results of tests other than the DIT were not available at the time of writing.

The experimental group pre- to posttest changes on the DIT were significant $t(26) = 2.37$, $p \leq .024$; the Lago placebo group showed non-significant changes at the .12 level; the second control group showed no significant changes. The Lago results, although not at an acceptable significance level $t(27) = 1.60$ suggest caution in interpreting the sources of the Sierra group's changes. Further analysis of between-group scores, such as analysis of variance would probably show no significant increases of the experimental group over the control.

Moral education programs.

Two of the studies with specific moral education content and focus failed to raise their subjects' moral judgment levels (Troth,1974; Siegal, 1974). Both were first-attempt studies on using their particular models, and involved untested assumptions about curriculum, design and evaluation manipulation. Troth's college "Values Course" was compared with a political science class. Both were intact classes, and neither yielded significant gains on the DIT over the experimental period. Troth's study is complicated in that she compared performance on only half the DIT in the pre-test with the full DIT as posttest.

Siegal's (1974) was a large-scale attempt (n = 350) to increase students' competence in evaluating moral issues and judgments. The design compared groups exposed to three modes of moral education and a control group. Group I was exposed to discussions of dilemmas generated by the experiments following Kohlberg's pattern. Group II was given the Meux packaged program, designed to train students to test their moral principles by increasing logical reasoning capacities. Group III was given the Association for Values Education and Research (AVER) program to foster social role-taking and logical reasoning in the realm of normative discourse. Group IV, the Inquiry/Control, group was trained in hypothesis testing. None of the treatments had been used previously by the experimenters and were being adapted while teachers were being trained in each of the techniques.

An analysis of covariance performed on the DIT posttest, with pre-test used as covariate showed no significant difference (p =.126), although the Kohlberg group has the highest means. Difference scores between pre- and posttest showed that the Meux treatment was relatively more effective in raising

DIT P score levels. Percentage increases (pre- to posttest) were Heux 12.3%, Kohlberg, 2.9%, AVER 1.3%, Con/Inquiry -1.1%. This raises some questions about the assumed superiority of discussion methods, since contamination across treatments may have contributed to the failure of the treatments to differentiate groups, or to achieve significant gain scores.

Piwko (1975) studied the effect of an experimental (E) Moral Development Workshop on 36 college freshmen. He used two control groups: one (C1) composed of 15 volunteers from a Human Development Course, and the second (C2) a random sample of 15 psychology students. The experimental course consisted of theoretical description of various approaches to moral development, exploration of moral implications of social issues, and the opportunity for subjects to explore their own personal values and value system. Only the experimental group showed significant upward movement from pre- to posttest $F(1,33) = 6.89$ $p < .05$. Since the treatment contained an explanation of Kohlberg's theory, as well as reflection on personal values, it is difficult to identify the developmentally influential variables in the workshop but the differences between it and the two control groups may be attributed, given sampling and analysis constraints, to some aspect of the course content, such as moral judgment experience, or acquaintance with the theory.

The only study outside formal education institutions was Coder's (1975) which used volunteer adult members from each of two liberal church congregations, Catholic church (group A), and a Congregational church (group B). A control Group C consisted of subjects enrolled in church seminars.

In addition to the DIT, the study used two of Rest's other moral judgment measures for pre- and posttests: the Moral Comprehension and the Law and Order Orientation tests, (Rest, Cooper, Coder, and Masanz, 1974). Liberalism and Radicalism-Conservatism scales, and SES and IQ tests were also used. Two

different moral education programs were compared. Group A followed the Kohlberg model of discussions of moral dilemmas, Group B received weekly lectures, with questions, but no discussion. DIT results showed no significant differences between any of the three groups on the pretest, and no differences between the two experimental groups on the posttest. On the posttest, the experimental groups, combined, differed from the control group, ($F=5.688$, $p \leq .005$). Scheffe and matched t test comparisons confirmed this difference at the .05 level.

The attribution of developmental change can be questioned on the ground that return by mail of the posttests stretched beyond the two week target period, to nine months in some cases. But nevertheless the significant difference between the experimental and treatment groups cannot be dismissed. Such variability presents some difficulties in interpreting the extent of the effects of the intervention treatments. However, the study, like Siegal's, did not confirm the widely-held assumption that discussions are more effective facilitators of moral judgment development.

Illustrative study: moral and judgmental treatment content.

Panowitsch (1975) illustrates the value of a moral education intervention built on the concept of the distinctive "moral" and "judgmental" dimensions of moral judgment, i.e., moral orientation, and judgmental activity. The DIT and Cornell Critical Thinking Test (CCTT) as tests of moral judgment and logical judgment respectively were used to compare students in four different college courses. The subjects were enrolled in Ethics and Logic (Experimental Groups), and World Religion and Art (Control Groups) classes over several school quarters.

One specific aim was to test the sensitivity of the DIT in differentiating gains in moral critical thinking (in the Ethics groups), from general critical thinking (in the Logic Groups), and general values content (in the World Religion and Art Groups). It was expected that if the DIT tested only conceptual

change, the ethics and logic groups would not differ in their posttest gains on the two instruments. If the DIT were simply a test of general values orientation, then the Ethics and the world religion groups would not differ on the DIT. But if the DIT were a sensitive test of moral judgment, then the Ethics groups would show significantly higher gains than the other two groups.

The study also provided opportunity to examine the content of a moral judgment intervention that used a treatment built on the concept that moral judgment-making involves solving moral problems in relation both to their content, and the problem-solver's judgmental processes. The Ethics courses had two components addressing this idea: a) reading and understanding the thought of moral philosophers, e.g., Kant, Neitzsche, and b) the application of methods of moral problem-solving to contemporary issues, e.g., suicide, abortion, civil disobedience. Students were given experience in understanding moral concepts, and in making their own moral judgments. In contrast, the Logic students were given training in formal logic, and logical problem-solving. World religion and Art students had experience with broad values issues.

Panowitsch's results confirmed hypotheses of the interrelationship between moral judgment theory, the ethics course, and the DIT, P index. The combined ethics groups showed significant gains on the DIT from pre- to posttest $\underline{t}(72) = 3.21$ $\underline{p} \leq .002$, while the logic and world religion and art classes showed no significant gains. Logic groups showed the only significant gains on the CCTT over the same length of time $\underline{t}(22) = 2.19$, $\underline{p} \leq .040$. The logic and ethics groups did not differ on the DIT pretest but were significantly different on the posttest $\underline{t} = 2.16$, $\underline{p} \leq .034$, (See Figure 1). Comparisons by the Scheffe test showed that the Religion and Art classes differed from the Ethics and Logic classes, but did not differ significantly from each other ($\underline{p} \leq .05$).

Panowitsch gave a five month follow-up test on the DIT to the students taking the Ethics and Logic courses in the Spring quarter. The changes for the Ethics subjects were significant for both pre- to posttest (p .003), and pretest to follow-up (p .015). From posttest to follow-up they did not change significantly (p .870) but showed a retention and plateau effect consistent with theory. The Logic subjects showed no significant gains from pre- to posttest, or from either pretest or posttest to follow-up. This study illustrates the value of: a) designing a moral education program in accord with the theoretical concepts of moral and judgmental components of moral thought, b) the DIT as a useful index of moral judgment for educational interventions, c) the use of a large sample, i.e., one hundred and fiftytwo students, and d) the addition of follow-up testing in a developmentally oriented study.

In that Panowitsch's subjects were not assigned to treatments randomly, but were in self-selected intact classes, there is a possible confounding of student interest and course content. To control for possible differential selection effects, Spring Ethics and Logic classes completed questionnaires asking students for their reasons for selecting either the ethics or the logic course. Reasons given for selecting the two courses were similar, and this was supported by absence of pretest differences between the two groups on the DIT and CCTT.

The conflation of several ethics and logic classes, across different instructors, into post hoc combined groups, presents further problems that the author acknowledges, and tried to deal with by several separate analyses (e.g., by age, sex, quarter), unfortunately with unequal sample sizes. On the whole, though, the results supported theoretical predictions, and the construct validity of the DIT.

In summary, eight of the fourteen studies, were successful in obtaining some sort of statistically significant change for experimental groups, on the DIT following intervention. Statistically significant results were more likely to be found in those studies where independent variables were rigorously specified & in which appropriate controls were used (See Table I).

DISCUSSION.

The results of these studies provide independent support for the ability of the DIT to measure upward movement in general psychological and moral judgment programs designed to produce developmental change. The meaning of the obtained gain-scores in relation to developmental and moral judgment concepts is discussed below. The instrument did differentiate between experimental and control groups, and between successful and unsuccessful treatments.

While no blueprint for the successful moral judgment intervention emerges, there are several indicators of what will contribute to confidence in the experimenter's claim to have facilitated change: the duration, design and characteristics of the experimental program. Walker's study is an example of the type that might well be dropped from moral education programs, i.e., the brief, one-lesson treatment. This kind of study assumes that structural and conceptual change can be produced instantly. Structural change involves the reorganization of the person's way of thinking about moral issues, and that

takes more than one whisk of the educational beater. All of the effective studies were of at least a quarter's duration. However, length of intervention, though it may be a necessary condition for effective intervention, is not a sufficient condition.

Design defects such as biased sampling and omission of control groups have especially severe effects in developmentally oriented studies where conclusions must be based on difference scores. The developmentalist must distinguish between artificially stimulated and "natural" developmental changes, and must be reasonably confident that obtained experimental changes do not reflect selection, maturation or historical effects, or some combination of these. In several of the studies, control groups as well as experimental groups showed considerable upward movement, complicating interpretation of the treatment effects (e.g., Hurt, Whiteley).

Sampling problems impose further constraints on interpretation. Within formal education institutions, random sampling can rarely be obtained for interventions (Snow, 1974). Sample selection, and assignment to groups are influenced by such factors as course enrollments, organization, time-tabling, etc. None of the experimenters drew random samples. Siegal assigned subjects to treatments randomly, Hurt assigned treatments randomly to intact groups, and Balfour's control group was chosen at random. Normally assignment to conditions was either by intact group, or volunteers for experimental programs. Such constraints require more conservative statistical analyses than the methods employed in the studies reported here.

Failure to control for "history" and "maturation" effects (Campbell and Stanley, 1963) are particularly damaging to inferences about sources of "developmental" change. The developmentalist needs adequate controls for effects between

pre- and posttestings, and for effects of time and stimulus variables on the internal processes of development within the subject.

Moral judgment treatments are not concerned with overt skill acquisition, but with changes in the subject's mode of thinking about moral issues. In order to infer that the treatment stimulated such change the experimenter must provide: adequate control groups; control of extraneous variables within the treatment group; description and observation of variables assumed to be developmentally critical; some monitoring of the assumed mechanisms of structural change; and developmentally appropriate testing.

While all the experimenters claimed some sort of developmental significance for their studies, two did not use control groups at all (Rest, Ahlgren, Mackey; and Erickson et al), and one only minimally (Sprinthall and Bernier). Several used levels of control (Hurt, Panowitsch, Whiteley), and others alternative moral treatments (Coder, Piwko, Siegal). Studies varied in the degree of control exercised over treatment variables and their relation to developmental or skill-related dependent variables. The reports of multitreatment studies make it difficult to know which of the numerous treatment factors actually influenced development. They ranged widely across such things as: teaching moral development theory, journal-keeping, empathy training, residential and community experience, and often seemed to be thrown in as a pot pourri. Erickson et al, Hurt, and Sprinthall and Bernier tried to monitor the effects of one skill, i.e., empathy training. Inclusion of auxiliary testing is an advance towards identification of effective moral judgment treatment variables.

Added to sampling and control constraints on the experimenter's claims, are statistical analyses and testing problems. Without randomization, the studies were dependent on pre- and posttest comparisons between experimental and control groups for some kind of equalization. Kickbusch and Siegal used analysis of covariance, with the pretest as the covariate, while the others mostly compared groups on

pretest measures. Only Coder, Morrison, Toews and Rest, Walker and Panowitsch compared experimental and control groups on the posttests. This omission of data in the other reports implies results may have been reported selectively. Most experimenters reported only pre- and posttest changes within treatment groups. Their claims that the treatment produced developmental movement are weakened by the omission of between group analyses.

Only Panowitsch and Morrison, Toews, and Rest used follow-up tests to measure enduring effects. Cognitive-development theory predicts either continued upward movement, or plateau effects. Testing for contrary-to-theory regression is also desirable. Wohlwill (1973) suggests that for conservation studies of conceptual change, immediate posttests may not be as informative as delayed posttests. The complexity of moral reasoning suggests that testing for persistence of changes is particularly important. Wohlwill also suggests that pretests of experimental groups, over a period, prior to the experiment could be made. This approach may avoid some of the problems of educational interventions and provide links between longitudinal and intervention studies: i.e., linking treatments to "natural developmental patterns."

In designing the intervention the experimenter has to be confident about the fit between his treatment and the dependent measure. The Minneapolis Police Report represents a class of study in which developmental change is expected as a by-product of the main focus, i.e., attitude change. Balfour and Morrison et al, are studies in which effects were expected to transfer from treatment variables assumed to relate to moral judgment. Such assumptions must be justified both theoretically and empirically. In multitreatment studies there is need for the generation of methods to measure effects of different aspects of the treatment, such as intraexperimental observation and testing (Snow, 1974) and correlations between skill and developmental measures.

Perhaps more important for cognitive-developmentalists is the need to devise ways of charting the internal mechanisms to which they ascribe explanation of structural and conceptual changes. Though not an easy task, this is long overdue. Flavell (1970) points out the difficulty of distinguishing "environmental" and "organismic" variables, i.e., those processes by which the child does something to the stimulus in-put. Within the field it has been common to invoke certain global, little explained organismic variables, like equilibration or moral conflict as the internal monitors of change. Yet no study has been found which charts the process of equilibration in moral judgment research, or describes experimentally how it works.

In this sense, there seems to have been little advance since Blatt's (1970) early interventions. Many studies are still built on the assumption that dilemma-related discussions, and "plus one modeling" induce moral conflict and stage-related movement. Yet these assumptions are without rigorous empirical testing, either by control of stimulus or microanalysis of the treatment's characteristics, and furthermore some studies suggest that other conditions are sometimes more effective, e.g. analysing moral problems (Panowitsch, Siegal) and lectures on moral issues (Coder).

GENERAL COMMENTS.

Apart from concerns for rigor and economical, efficacious and appropriate use of program resources, two important questions are raised by such a review as this: "What causes an experimenter to adopt a developmental perspective for his study?" And, "Why choose moral judgment development as the dependent variable?

Why adopt a developmental perspective?

Invariably moral and psychological education programs are claimed to have developmental significance. Yet it seems that experimenters do not weigh the values and costs of adopting a developmental mode. An obvious advantage is

the purchase of contextual meaning for the results, both in relation to the subject's life-span, and in plugging into a body of research on developmental trends. It is easy to tread the well-worn pathway, gain contextual meaning, and fail to identify developmentally influential variables. Thus unmonitored treatments, or those invoking unexplained mechanisms, may proliferate reports, but they do not advance knowledge of moral judgment development.

Yet there are costs to exchanging isolated interpretation for the fellowship of the developmentally oriented. There are some major problems inherent in developmental research. For instance, how can gain scores on a particular index be interpreted in terms of actual growth in moral reasoning? There is no firm ground for assuming that scores on our rather crude measures reliably mirror changes in subject's mental operations. To measure moral judgment change the experimenter has to use either a stage-related index, or a single continuum score like the P and D indices. A predominant-stage score loses much vital information. The P and D indices, even though reliable and sensitive to movement, still measure only rather gross, end-product aspects of moral judgment. The field needs fine-grained methods of analysis, especially for before-and after-treatment assessment. Rest (1976a, 1977) has suggested the use of a profile of the individual subject's distribution of scores over all stages. This would avoid collapsing data to a single number. It would also permit the closer matching of treatment to a subject's individual level of responding to moral problems, that is, of course, if appropriately varied treatments can be developed.

Even if moral judgment levels could be accurately diagnosed and measured, interpretation problems are compounded by situational, personal, and interactive effects, and the test-taking factors that can intrude between the person's moral reasoning and test scores. Change scores may simply reflect either the

subject's increased understanding of the test task, or good feelings about the experience. Paradoxes abound in the test-taking complexities of such distinctions as competence and performance, structure and content; and hypothetical test and real-life moral problems.

Why moral judgment as the dependent variable?

If the developmental perspective is suitable, or chosen, for the study, why moral judgment development? Some of the enthusiasm may be due to the DIT's ease of administration and analysis. It can be given as a paper and pencil test in a group situation, and is computer analyzed. Another factor is the scarcity of developmental measures. The DPE model tries to build up a composite picture of general conceptual growth, and includes the DIT in the dependent variables. Yet it seems there ought to be a rationale for using a measure which is specifically "moral" and "Judgmental" in content. Results show that the DIT indexes some sort of change in general psychological development programs, and furthermore, correlational studies comparing it with other developmental maasures may confirm that it is as useful as some cognitive and ego development indexes of general conceptual change. But its purpose is to index one particular aspect of human thought and morality: the moral judgment. Panowitsch's study shows its ability to do this.

This review has shown the relative uncertainty with which moral educators can base their programming on research. Lockwood's review of Kohlbergian studies (1977) also questions the strength of current models of moral development interventions. Both the Siegal and the Coder results raise serious questions about the dilemma-cum-discussion method for moral education programs. Coder's lecture treatment was as effective as the standard Kohlbergian approach. Siegal's adaptation of the discussion method was no more successful than the attempt to

influence upward movement by developing subjects' moral reasoning skills. Little is known about effective environmental variables, especially in multi-treatment studies. Still less is known about organismic processes of change. There is need for more experimental investigation of the twin concepts of moral criteria and judgmental skills.

DIRECTIONS FOR FURTHER RESEARCH.

How can moral judgment intervention studies be made more developmentally informative? Apart from adherence to sampling, control and testing standards, objectives need to be made explicit in relation to moral judgment theory, and more accurate diagnostic measures need to be devised. If the DIT is used, it should be for its theoretical assumptions, and their suitability to the independent variables. In-put variables believed to be developmentally influential need to be monitored and measured: a) intraexperimentally, b) in relation to the developmental dependent variables, and c) for cumulative, generalizable, and continuing, i.e., developmental, outcomes over time, in follow-up tests. If listening skills, discussion participation, or log-keeping are assumed to stimulate growth, then the experimenter needs to report, for example, how much a subject took part in discussion, and how his performance correlates with DIT scores, and how it compares with that of the non-participant. This is one way of linking critical variables to developmental outcomes.

Finally the study should be placed within a research program of replication and refinement. A number of the studies reported here were either dissertation studies or the experimenter's first attempt at operationalizing some hunches, or a particular program. This factor in itself places certain constraints on what can be achieved. In the first run, the experimenter often has limited control over developmental variables, curriculum materials, teacher

training or the sometimes strange climate of the classroom, as well as experimental and design factors. The first attempt should not ordinarily be expected to validate procedures for theoretical presuppositions, but rather, to link present research to past findings, and to identify variables of developmental interest.

Placed within an ongoing program, the findings of the study can feed back into the next wave of research. Transition studies should be used to sort out effective variables from "multitreatment interference" (Bracht and Glass, 1968). In summary three specific suggestions emerge from this review: interventions should be tailored to fit moral judgment theoretical concepts, the measurement instrument should be a good fit for the treatment, and future research should be focused on moral and judgmental aspects of moral judgment programs.

References

Allen, R. An evaluation of the Nicolet High School Confluent Education Project for the school year 1974-1975. Unpublished manuscript, 1976.

Balfour, M. J. An investigation of a school-community involvement program's effect on the moral development of its participants. Unpublished Master of Arts dissertation, University of Minnesota, 1975.

Blatt, M. Studies on the effects of classroom discussion upon children's moral development. Unpublished doctoral dissertation, University of Chicago, 1970.

Bracht, G. H. and Glass, G.V. The external validity of experiments. American Educational Research Journal, 1968, 5, 437-474.

Campbell, D.T. and Stanley, J. C. Experimental and Quasi-experimental research on teaching. In N. L. Gage, Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 171-246 and Reprinted R. McN., 1966.

Coder, R. Moral judgment in adults. Unpublished doctoral dissertation, University of Minnesota, 1975.

Davison, M. L., Robbins, S. and Swanson, D. Stage structure in objective moral judgments. Submitted for publication 1977, in press.

Erickson, V. L., Colb , S., Libbey, P., and Lohman, G. The young adolescent: A curriculum to promote psychological growth. In G.D. Miller (Ed), Developmental Education and Other Emerging Alternatives in Secondary Guidance Programs. Minneapolis: Minnesota Department of Education, 1976.

Flavell, J. H. "Concept Development" In P. Mussen (Ed), Carmichael's Manual of Child Psychology. Vol. I. New York: John Wiley and Sons, Inc., 1970, 983-1059.

Hurt, B. L. Psychological education for teacher education students: a cognitive-developmental curriculum. Paper presented at the annual convention of the American Psychological Association, September, 1975, Chicago.

Lockwood, A. L. The effects of values clarification and moral development curricul on school-age subjects: a critical review of recent research. Unpublished manuscript, University of Wisconsin, 1977.

Morrison, T. R., Toews, B., Rest, J.R. Youth, Law and Morality. Unpublished manuscript, Winnipeg, 1975.

Nelson, K. H. The Sierra project: from theory to research. Paper presented to the Moral Development and Psychological Education Conferences, 1976.

Panowitsch, H. R. Change and stability in the Defining Issues Test: an objective measure of moral development. Unpublished doctoral dissertation, Uni ersity of Minnesota, 1975.

Piwko, J. The effects of a moral development workshop. Unpublished manuscript, 1975.

Rest, J.R. A theoretical analysis of moral judgment development. (Manuscript in preparation for publication) University of Minnesota, 1976(a).

Rest, J.R. Manual for the Defining Issues Test: An Objective Test of moral judgment. University of Minnesota, 1974(a).

Rest, J. R. New approaches in the assessment of moral judgment. In T. Likona, (Ed.), Moral Development and Behavior: Theory and Research and Social Issues. New York: Holt, R. and Winston, 1976(b), 198-220.

Rest, J. R. The stage concept in moral judgment research. Unpublished manuscript, University of Minnesota, 1977.

Rest, J.R., Ahlgren, C., Mackey, J. Minneapolis Police Report. Unpublished manuscript, University of Minnesota, 1972.

Rest, J.R., Casper, D., Coder, R., Masanz, J., Anderson, D. Judging the important issues in moral dilemmas - an objective measure of development. Developmental Psychology, 1974, 10 (4), 491-501.

Siegal, M. An experiment in moral education AVER in Surrey. Paper presented at Ann. Conf. Canada Society for the Study of Education, Toronto, Canada, June 7, 1974.

Snow, R. Representative & Quasi - Representative Designs for research on teaching. Review of Educational Research, 1974, Vol. 44, No. 3, 265-292.

Sprinthall, N.A. and Bernier, J. E. Moral and cognitive development for teachers: a neglected area. Chapter for Fordham University Symposium: Programs and Rationale in Value - Moral Education, 1977.

Tracy, J. J. and Cross, H.J. Antecedents of shift in moral judgment. J. of Personality and Social Psychology, 1973, 26, 238-244.

Troth, A.G. An assessment of impacts on students of the "Values" course. Unpublished manuscript, St. Olafs College, Northfield, 1974.

Turiel, E. An experimental test of the sequentiality of developmental stages in the child's moral judgments. Journal of Personality and Social Psychology, 1966, Vol. 3, No. 6, 611-618.

Walker, L. The effects of a narrative model on stages of moral development. Unpublished manuscript, University of New Brunswick, 1974.

Whiteley, J. M. The scope and purpose of the UCI residential education project. Unpublished manuscript, University of California, Irvine, 1976.

Wohlwill, J.F. The study of Behavioral Development. New York: Academic Press, 1973.

TABLE I. OVERVIEW OF INTERVENTION STUDIES, USING THE DIT.

Jeanette Lawrence, 330 Burton Hall
University of Minnesota

Study	Sample	Type of Intervention	Duration	Design	Change on DIT	Characteristics
SHORT-TERM STUDY:						
1. Walker, 1974	8th grade N=70	narrative modeling of pro, con reasoning (+1), (-1), own stage	one-lesson treatment	3 exp. groups (+1), (-1), (0)	Es: no change	assumes immediate, easy change and stage typing
SOCIAL STUDIES PROGRAMS:						
2. Mpls. Police Report, 1972	Jr. high N=61	soc. studies instructional unit to change attitudes to police	12 weeks	exp. groups no control	E: no change on P, decrease in use of St.2, t=-2.17, p<.05	attitude change study, with developmental measure, no control group
3. Morrison, Toews, Rest, 1975	Jr. high N=103	"Youth, Law and Morality" Program: social & civics studies, to develop moral reasoning	4 months	exp. group control groups (follow-up test six weeks)	Es: no change Cs: no change	multivariable treatment use of follow-up test
PSYCHOLOGICAL DEVELOPMENT PROGRAMS:						
4. Erickson, Colby, Libbey, Lohman, 1975	Jr. high N=20	DPE Curriculum for personal development, within school classes	semester	exp. group, no control	E: t (19) = 2.27, $p \leq .02$	auxiliary empathy scale taught Kohlberg stages (no control group)
5. Hurt, 1975	College N=54	counselling, empathy training, in educational psychology course	quarter	exp. group, (E) active con. (C1) inactive con(C2)	E: $t(14)=1.94$, $p\leq.037$, C1: $t(18) = 1.80$, $p\leq.045$, C2: NS	treatments randomly assigned to intact groups, aux.testing, 13 inconsistents in E. group
6. Sprinthall Bernier, 1977	in-service teachers N=18	intensive workshop in personal, professional development, plus seminar while teaching	6 weeks workshop, + qrter.	exp. group	E: $t(17)=2.91$, $p<.01$	taught theory, skill index aux. testing, comparison group to relation to E
7. Balfour, 1974	Sr.high N=84	"Humanities Outreach Course": community experiences, seminar	semester	3 exp. groups, in 3 schools, 1 control	Comb.E: $t(53) = 2.01$, p .05; West.E: $t(11)=1.83$, $p\leq.046$ C: no change	multivariate treatment, transfer to DIT D.V.
8. Kickbusch, Allen, 1976	9th grade N=117	"Confluent Education" for integration of affective, cognitive aspects of curriculum teaching style, unit on morality	8 months with semest. in moral educ. unit	exp. groups (E) Control (C1) Control (C2)	E: no change C1: no change C2: no change	multiplicity of treatment variables and testing, inconsistency norm on DIT

Study	Sample	Type of Intervention	Duration	Design	Change on DIT	Characteristics
9. Sierra Project, Whiteley, Nelsen, 1976	College freshmen N=77	Residential, community living program to integrate formal and informal education	8 months, with moral education courses	exp. group, placebo con(C1) control (C2)	E: $t(34)=2.37$, $p \leq .024$, C1: $t(26) = 1.6$, $p < .12$(NS), C2: no change	multiple experiences, course module on moral dev. theory. Ran ass. of volunteers to E. and C1, ran sample C2
MORAL EDUCATION PROGRAMS:						
10. Troth, 1974	College N=42	"Values Course": to integrated personal values and behavior	semester	exp. and control groups	E: no change C: no change	pretest (1/2 DIT), posttest (whole), new program
11. Siegal, 1974	8th, 9th 10th grade N=358	Treatments: I. Kohlberg, II. Meux, III. Aver, IV. Inquiry/control	semester	3 exp. groups, 1 control	E: no change C: no change	large scale, confounding of treatments, materials, and teacher training
12. Piwko, 1975	College freshmen N=36	"Moral Development Workshop": moral values, commitment, Human development course (C1) no treatment (C2)	quarter	ext. groups, 2 controls	E: $F(1,33)=6.89$, $p < .05$, Cs: no change	course focused on exploration of moral issues
13. Coder, 1975	adults: church members N=87	A. Dilemma and discussion, B. Lecture, no discussion C. Other seminars	6 weeks, (x 2 hrs.)	2 exp. groups, (A and B) 1 Control (C)	(A+B) differs from C on Posttest: $F=5.69$, $p < .005$ NS diffs. between A & B	discussion lecture, (not significantly different) delay in return of posttests
ILLUSTRATIVE STUDY:						
14. Panowitsch, 1975	College N=152	E: Ethic class, L: Logic class, World religions class (C1) Art class (C2) comparison of moral judgement, critical thinking, and general values treatments	quarter + 5 months follow-up for 2 sub-samples	2 exp. groups, 2 controls (follow-up 8 months)	Comb. E: pre-post DIT $t(72)=3.21$, $p \leq .002$, Comb. L: pre-post CCTT, $t(22) = -2.91$, $p \leq .040$, Controls NS, E vs L: pre-NS, post- $t = 2.16$, $p \leq .034$ E: post- to follow-up, NS.	large sample, levels of control, use of follow-up, operationalized "moral", "judgmental" content, intact classes.